



Overdiagnosis of lung cancer with low-dose computed tomography screening meta-analysis of the randomised clinical trials

Brodersen, John; Voss, Theis; Martiny, Frederik; Siersma, Volkert; Barratt, Alexandra;
Heleno, Bruno

Published in:
Breathe

DOI:
[10.1183/20734735.0013-2020](https://doi.org/10.1183/20734735.0013-2020)

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Brodersen, J., Voss, T., Martiny, F., Siersma, V., Barratt, A., & Heleno, B. (2020). Overdiagnosis of lung cancer with low-dose computed tomography screening: meta-analysis of the randomised clinical trials. *Breathe*, 16(1), [200013]. <https://doi.org/10.1183/20734735.0013-2020>



Credit: Prtump16 via WikiMedia Commons, CC-BY-SA 4.0

Key points

- Nine randomised controlled trials (RCTs) on low-dose computed tomography screening were identified; five were included for meta-analysis but only two of those were at low risk of bias.
- In a meta-analysis of recent low risk of bias RCTs including 8156 healthy current or former smokers from developed countries, we found that 49% of the screen-detected cancers may be overdiagnosed.
- There is uncertainty about the degree of overdiagnosis in lung cancer screening due to unexplained heterogeneity and low precision of the point estimate.
- If only high-quality RCTs are included in the meta-analysis, the degree of overdiagnosis is substantial.

Educational aims

- To appreciate that low-dose computed tomography screening for lung cancer meets all three main conditions for overdiagnosis in cancer screening: a reservoir of indolent cancers exists in the population; the screening test is able to “tap” this reservoir by detecting biologically indolent cancers as well as biologically important cancers; and the population being screened is characterised by a relatively high competing risk of death from other causes
- To learn about biases that might affect the estimates of overdiagnosis in randomised controlled trials in cancer screening

Review

Overdiagnosis of lung cancer with low-dose computed tomography screening: meta-analysis of the randomised clinical trials

In low-dose computed tomography (LDCT) screening for lung cancer, all three main conditions for overdiagnosis in cancer screening are present: 1) a reservoir of slowly or nongrowing lung cancer exists; 2) LDCT is a high-resolution imaging technology with the potential to identify this reservoir; and 3) eligible screening participants have a high risk of dying from causes other than lung cancer. The degree of overdiagnosis in cancer screening is most validly estimated in high-quality randomised controlled trials (RCTs), with enough follow-up time after the end of screening to avoid lead-time bias and without contamination of the control group.

Nine RCTs investigating LDCT screening were identified. Two RCTs were excluded because lung cancer incidence after the end of screening was not published. Two other RCTs using active comparators were also excluded. Therefore, five RCTs were included: two trials were at low risk of bias, two of some concern and one at high risk of bias. In a meta-analysis of the two low risk of bias RCTs including 8156 healthy current or former smokers, 49% of the screen-detected cancers were overdiagnosed. There is uncertainty about this substantial degree of overdiagnosis due to unexplained heterogeneity and low precision of the summed estimate across the two trials.

Cite as: Brodersen J, Voss T, Martiny F, *et al.* Overdiagnosis of lung cancer with low-dose computed tomography screening: meta-analysis of the randomised clinical trials. *Breathe* 2020; 16: 200013.

Broadly, overdiagnosis means making people patients unnecessarily, by identifying problems that were never going to cause harm or by [over]medicalising ordinary life experiences through expanded definitions of diseases.

BRODERSEN *et al.* [1]

Overdiagnosis has two major causes: overdetected and overdefinition. In cancer screening, it is overdetected that is in play and it refers to identification of indolent cancer pathology: pathology that will never cause harm, never progress,

progress too slowly to cause symptoms or harm during a person's remaining lifetime, or that resolves spontaneously [2]. Many countries are currently considering the introduction of low-dose computed tomography (LDCT) screening for lung cancer in high-risk groups. Good policy decisions about screening are informed by the best available evidence concerning the benefits, harms, costs and ethical implications of screening for individuals and for society. In this regard, the extent of overdiagnosis due to lung cancer screening requires attention as it is the most severe potential harm of screening and because LDCT screening for lung

 @ERSpublications

In low-dose computed tomography (LDCT) screening for lung cancer, all three main reasons for overdiagnosis are present. Half of people detected as having lung cancer *via* their participation in LDCT screening are overdiagnosed. <http://bit.ly/32tLZk4>



CrossMark



© ERS 2020

cancer demonstrates all three main conditions for overdiagnosis due to overdetection [2, 3]:

- there is empirical evidence of substantial heterogeneity in growth rates of LDCT screening-detected lung cancers, indicating that a reservoir of slowly or nongrowing lung cancer exists [4–6].
- current and former heavy smokers have a high risk of dying from causes other than lung cancer; for example, from cardiovascular diseases.
- LDCT scans have a much higher resolution than chest radiography, thus increasing its ability to detect the reservoir of indolent/slow-growing pathology.

Screening with chest radiography has previously been shown to result in overdiagnosis of lung cancer [7]. Therefore, LDCT screening for lung cancer will inevitably result in some degree of overdiagnosis. Because LDCT can detect smaller cancers than can chest radiography, CT scanning would be expected to find more cancers, and the absolute number of both biologically important and of biologically indolent cancers will be increased. Furthermore, if even smaller lung cancer nodules have even slower growth rates than the bigger nodules – or no growth at all – then the degree of overdiagnosis in CT screening would theoretically be expected to be greater compared to the degree of overdiagnosis in chest radiography [2].

Overdiagnosis may result in all the harms of regular diagnosis, from which it is indiscernible, but, by definition, without providing any benefits. Harms range from physical harm from unnecessary diagnostic tests and treatments; to psychosocial harm when people are told that they have a fatal disease; to economic harm, such as loss of income while undergoing tests and treatment; to opportunity costs, such as time that could be otherwise spent with loved ones; and finally, to societal costs, since the resources used in screening and downstream procedures could be used in other healthcare activities [8]. Moreover, overdiagnosis occurs to a greater or lesser extent in screening programmes that target the actual cancer and not its precursors [9]. So, to balance the potential intended benefits and the inevitable unintended harms of cancer screening, an assessment of overdiagnosis is needed [10, 11].

Overdiagnosis is, together with false positives and a general increased fear in society, an important driver of harms of screening [12–14]; thus, it should be reported in every trial assessing screening. The degree of overdiagnosis in cancer screening is most validly estimated in high-quality randomised trials with enough follow-up time to avoid lead-time bias and with no contamination of the control group [3, 15]. Lead time is the length of time between screen-detected lung cancer and the theoretical time point of the clinical presentation of lung cancer in a setting with no screening. In the present paper,

sufficient follow-up time to avoid lead-time bias is 3.6 years [16]. Contamination is defined as the proportion of the control group who have been screened; the higher the contamination rate, the more the estimate of overdiagnosis is biased towards the null.

For example, the US National Lung Cancer Screening Trial (NLST) compared three rounds of screening with LDCT (n=26 722) to three rounds of screening with chest radiography (n=26 730) [17]. After a mean follow-up time of 6.4 years, 18.5% of the screen-detected cancers were estimated to be overdiagnosed [18]. The follow-up time is enough to account for the lead-time, *i.e.* to ensure that all cancer that would have presented clinically have been diagnosed in the control group. However, as screening with chest radiography has also been shown to result in overdiagnosis of lung cancer [7], the control group is contaminated and the degree of overdiagnosis for LDCT screening is underestimated in the NLST.

In view of the contamination due to using active comparators in screening trials, the aim of this study was to estimate the degree of lung cancer overdiagnosis due to LSCT screening compared with no screening. To provide the best quality evidence, we performed a meta-analysis of randomised controlled trials (RCTs) only.

Methods

Eligibility criteria

Randomised trials were eligible if they reported the incidence of lung cancer for people screened with LDCT compared to people who were not screened (usual care).

Randomised trials were excluded if:

- they did not provide long-term cumulative lung cancer incidence during follow-up, *i.e.* after the active phase of trials; or
- the control group was offered any type of lung cancer screening after or during the RCT.

In all but one trial, screening participants were at high risk of lung cancer due to being either former or current smokers with a history of ≥ 20 pack-years. In the Chinese Yang trial, participants were also eligible if they had a family or personal cancer history, occupational exposure, second-hand smoking or long-term exposure to cooking oil fumes. We included trials regardless of risk of bias. There were no restrictions concerning date of publication or language.

Search strategy

We searched PubMed for reviews of lung cancer screening with LDCT using the search terms “screening”, “low-dose computed tomography” and

“lung cancer”, or the name of the trials we knew of that had evaluated or are evaluating lung cancer screening with LDCT, in the title, abstract or keywords of publications. We extracted the references to lung cancer screening trials included in these reviews. From our >15 years of research in the Danish Lung Cancer Screening Trial (DLCST), we knew of eight RCTs comparing LDCT to no screening [19]:

- Chinese Yang Trial
- DLCST [20]
- German Lung Cancer Screening Intervention Study (LUSI) [21]
- Multi-Centric Italian Lung Detection Trial (MILD) [22]
- Italian Lung Cancer Computed Tomography Screening Trial (ITALUNG) [23]
- Detection and Screening of Early Lung Cancer by Novel Imaging Technology and Molecular Assays (DANTE) [24]
- Nederlands-Leuvens Longkanker Screenings Onderzoek (NELSON) [25]
- UK Lung Cancer Screening trial (UKLS) [26]

One more RCT on LDCT screening we knew of was the American NLST [17], which compared LDCT to chest radiography. The NLST was not included in the present review due to the active comparator.

Data collection and extraction

Selection of studies

If any of the eight trials were reported in multiple articles, we selected the article reporting the longest follow-up time.

Data extraction and management

Two authors (T. Voss and B. Heleno) independently extracted data from the included trials and entered them into a data extraction form in Excel (Microsoft Corp., Redmond, WA, USA). Disagreements were resolved through discussion with a third review author.

We extracted the following data:

- Inclusion and exclusion criteria for the study population
- Characteristics of participants (sex, age and number enrolled in each group)
- Number of screening rounds and interval between screening rounds
- Characteristics of any comparator screening intervention
- Definition of an abnormal screening result
- The incidence of cancer in the screened group and in the control group
- Number or proportion of screen-detected cancers in the screened group
- The participation rate (participation defined as having had at least one screening test)

- The contamination rate (number in the control group with at least one screening test during the active phase or during follow-up)
- The rate of screening in both arms of the trial after the active phase
- The duration of follow-up

Assessment of risk of bias

The Cochrane Risk of Bias tool

Two authors (T. Voss and B. Heleno) independently assessed the risk of bias of the included RCTs using the Cochrane Collaboration's Risk of Bias tool 2.0 [27], which includes the following domains.

- Risk of bias arising from the randomisation process.
- Risk of bias due to deviations from intended interventions. Contamination bias was assessed under this dimension.
- Risk of bias due to missing outcome data.
- Risk of bias due to measurement of the outcome. Lead time bias was assessed under this dimension. As mentioned in the introduction, we considered that trials adequately accounted for lead-time when follow-up since the last screening round was >3.6 years [16].
- Risk of bias due to selection of the reported result.
- Overall risk of bias assessment.

Data management and statistical analysis

To estimate the degree of overdiagnosis, we calculated the risk ratio (RR) of lung cancer and the respective 95% confidence intervals. The RR expresses the increase or the decrease in the incidence of lung cancer at the population level if LDCT screening for lung cancer screening was to be implemented. If the lung cancer incidence in the screened group and the lung cancer incidence in the control group are the same, with a sufficiently long follow-up time, then there is no overdiagnosis. If the lung cancer incidence in the screened group is greater than the lung cancer incidence in the control group, again with a sufficiently long follow-up time, this is evidence of overdiagnosis. And finally, if the lung cancer incidence in the screened group is smaller than the lung cancer incidence in the control group, also under the condition of a sufficiently long follow-up time, then the screening prevents cancers from occurring (the screening has a primary preventive effect).

We also calculated another measure of overdiagnosis that has been used in several previous studies and thereby can be used for comparisons representing the risk that a screen-detected cancer is overdiagnosed. This standard measure is more informative when an individual eligible screening participant is to be informed about pros and cons of screening, *i.e.* what is the likelihood that the cancer

is overdiagnosed should you be detected with a cancer in the screening programme [3, 28]:

$$\left(\frac{\text{Cumulative cancer incidence in the screened population}}{\text{Number of subjects screened}} - \frac{\text{Cumulative cancer incidence in the control population}}{\text{Number of control subjects}} \right) \times 100$$

$$\left(\frac{\text{Cumulative number of screen-detected cancers}}{\text{Number of subjects screened}} \right)$$

The risk that a screen-detected cancer is overdiagnosed is more sensitive to the baseline incidence of cancer diagnosis compared to the RR of cancer. Therefore, we calculated both measures. The cumulative cancer incidence in the screened population was defined as all cancers detected in the population offered screening during and after the active phase. Likewise, the cumulative cancer incidence in the control population was defined as all cancers detected in the control population during and after the active phase. The cumulative number of screen-detected cancers was defined as all cancers detected by screening in the population offered screening during the active phase.

Using R (The R Foundation, Vienna, Austria), we calculated standard deviations of this measure through bootstrapping and used a normal approximation to compute 95% confidence intervals.

We assumed that the baseline risk of lung cancer in the absence of screening was the median incidence of lung cancer in the control arms of the included trials.

Data synthesis

We pooled overdiagnosis estimates across included RCTs. Results were summarised with a random-effects meta-analysis using the inverse-variance method, as we anticipated some variation due to the different timing, follow-up and settings of individual trials. Data were analysed using Review Manager version 5.3 (The Nordic Cochrane Collaboration, Copenhagen, Denmark).

Sensitivity analysis

We performed a secondary meta-analysis restricted to trials with low risk of bias across all domains.

Results

In our literature search, we identified the nine trials investigating LDCT screening, of which we already knew (table S1) [17, 23–26, 29–32]. This paper was first submitted to the journal *Breathe* on 20 January 2020. On 29 January 2020, the NELSON trial published its results [33], and we included the

NELSON trial's results in a revised version of the manuscript. Therefore, five trials were included in the present study: DLCST, ITALUNG, LUSI, MILD and NELSON (table 1) [23, 29–31, 33]. Two trials (UKLS and the Chinese Yang trial) were not included since no data on the incidence after the end of screening in these two trials were published [26, 32]. Two trials used active comparators and therefore were not included: the NLST compared LDCT screening with chest radiography and DANTE offered a single baseline chest radiography screening to the control group [17, 24].

The main characteristics of each trial and our risk of bias rating are summarised in table 1. The median cumulative risk of lung cancer in the control groups was 3.96% during the active screening phase in the intervention groups across the five trials. Two of the included trials (DLCST and LUSI), accounting for 21.6% of the lung cancers, were at low risk of bias [29, 30]. There were some concerns regarding the ITALUNG trial, since the extent of contamination in the control group was not reported [23]. The MILD trial was at high risk of bias because of baseline differences (89% current smokers in control group *versus* 68.3%/68.8% in intervention group, and it only started randomising participants to an unscreened control group 6 months after it started randomising participants). Recent results from the NELSON trial focus on men even though the protocol of the NELSON trial and earlier publications suggest that data about men and women would be reported in the same analyses [25]. Moreover, data about contamination in the NELSON is restricted to the baseline round [34].

Main analysis

We found that LDCT screening of former or current smokers increases the cumulative incidence of lung cancer (RR 1.22) with a 95% confidence interval of 1.02–1.47 and heterogeneity (I^2) 55% (figure 1). Of the screen-detected cancers, we estimate that 38% (95% CI 14%–63%) may be overdiagnosed (I^2 =65%) (figure 2).

Sensitivity analysis

Findings from the sensitivity analysis restricted to the two trials at low risk of bias in all dimensions suggests a larger RR point estimate (1.51, 95% CI 1.06–2.14; I^2 =58%) (figure 3) and similar probability that a screen-detected cancer is overdiagnosed (49%, 95% CI 11%–87%) (figure 4).

Discussion

Summary of main findings

Using data from the latest publications of low risk of bias trials comparing screening by LDCT with

Table 1 *Characteristics of included studies*

Trial name	Population characteristics	Intervention (screening interval months/rounds)	Total subjects (intervention/control arm subjects)	Control	Follow-up time after last screen years	Participation rate	Contamination rate
DLCST	Men and women aged 50–70 years, smokers and former smokers ≥20 pack-years Former smokers must have quit <10 years prior	LDCT (12/5)	4104 (2052/2052)	Usual care	5	95.5%	20.3%
ITALUNG	Men and women aged 55–69 years, smokers and former smokers ≥20 pack-years in past 10 years	LDCT (12/4)	3206 (1613/1593)	Usual care	5	81%	Not reported
LUSI	Men and woman aged 50–69 years, smokers and former smokers (cessation <10 years) with ≥25 years of smoking ≥15 cigarettes a day or 30 years of smoking 10 cigarettes per day	LDCT (12/5)	4052 (2029/2023)	Usual care	3	>90%	8.7%
MILD	Men and women aged 49–75 years, current or former (cessation <10 years) ≥20 pack-years	LDCT (12 or 24/6 or 3) Median duration of screening was 6.2 years	4099 (2376/1723)	Usual care	The follow-up since last screening round is unclear	95.1% in the biennial and 96.1% in the annual LDCT group	1.2%
NELSON	Men aged 50–74 years, current or former smokers (who had quit ≤10 years ago) who had smoked >15 cigarettes a day for >25 years or >10 cigarettes a day for >30 years	LDCT (12, 24, 30/4)	13195 (6583/6612)	Usual care	4.5	85.8% in total (lowest at round 4 with 67.4% and highest at round 1 with 95.8%)	Not reported

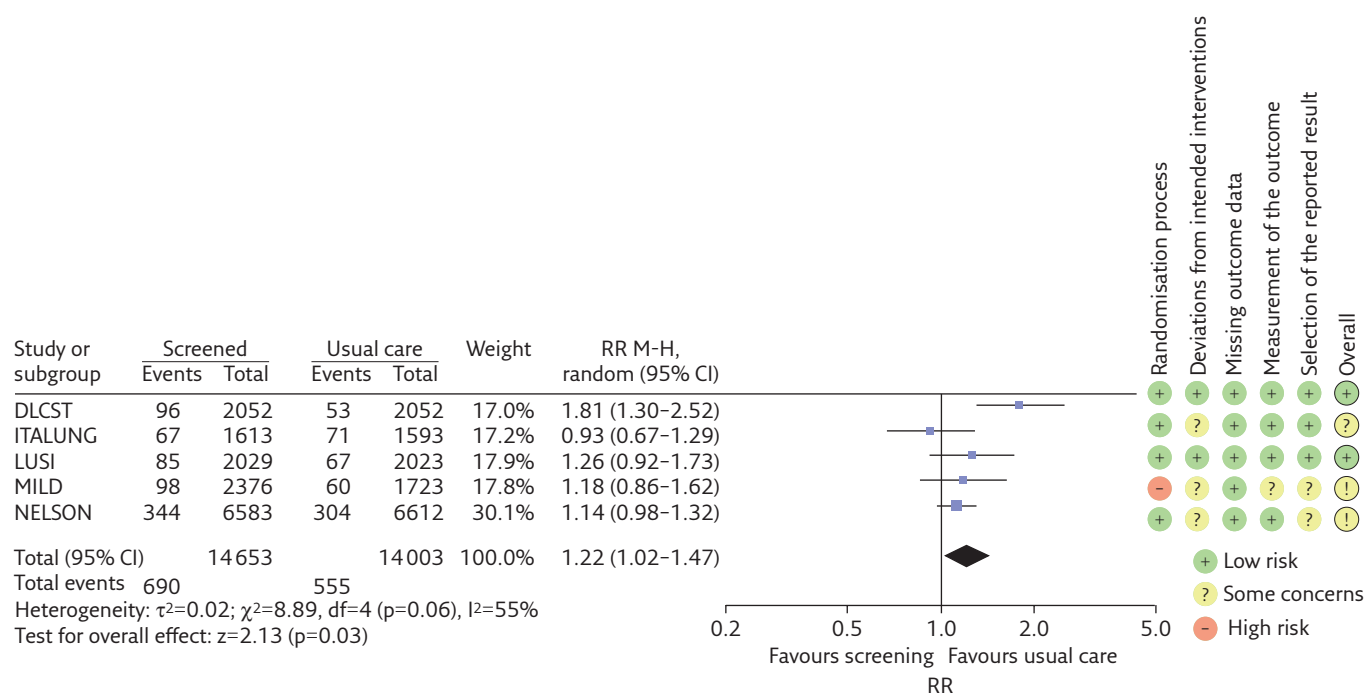


Figure 1 Forest plot of the RR of the cumulative incidence of lung cancer (estimates >1 represent overdiagnosis). The meta-analysis includes all trials, regardless of bias assessment. Trials are listed alphabetically. M-H: Mantel-Haenszel; df: degrees of freedom.

usual care, we found that screening annually five times with 3–5 years of follow-up increased the cumulative incidence of lung cancer (1.51, 95% CI 1.06–2.14). This corresponds to statistically significant absolute risk increase of 20 cancers per 1000 people screened, assuming a median baseline risk in the control group of 40 lung cancers per 1000 participants. We estimated that 49% (95% CI 11–87%) of the LDCT screen-detected cancers may be overdiagnosed. Using data from all included trials, regardless of the risk of bias, we found an increased cumulative incidence of lung cancer (RR 1.22, 95% CI 1.02–1.47), which corresponds to an absolute risk increase of nine cancers per 1000 people screened. Here, we estimated that 38% (95% CI 14–63%) of the LDCT screen-detected cancers may be overdiagnosed.

Strengths and weaknesses

We used prior reviews of lung cancer screening to identify trials of LDCT screening and selected those that had an unscreened control group with published data on long-term cumulative lung cancer incidence. One potential weakness of our approach is that our meta-analysis is based on a rapid review, which might lead to unintentional omission of important trials and/or publications. However, we consider it unlikely that we have missed important trials and/or publications concerning LDCT screening for lung cancer due to our commitment to this field for >15 years. Similar to systematic reviews, we have extracted key data from all included trials, presented in table 1, and we have performed a thorough risk of bias assessment of

included trials, independently by two review authors. Therefore, we believe that we have accounted for key differences between included trials.

The quality of the primary studies included in the meta-analysis was moderate. Two trials (DLCST and LUSI), accounting for around 22% of the lung cancers included in the meta-analysis, were at low risk of bias. The major threats to the confidence in our estimates are the high, unexplained heterogeneity between all five included trials ($I^2=55\%$) and the imprecision of the pooled estimate (the confidence interval for the meta-analysis spans 1.02–1.47). The heterogeneity was still high ($I^2=58\%$) in our sensitivity analysis of low risk of bias trials and consequently, the precision was low (1.06–2.14).

If we assume a baseline risk of 40 lung cancers per 1000 participants (corresponding to the 2.58% median cumulative risk of lung cancer in the control groups across included trials), our results in the low risk of bias trials are compatible with an absolute risk increase between two cancers per 1000 (based on the lower bound of the confidence interval of 1.06) and 46 per 1000 participants screened (based on the upper bound of the confidence interval of 2.14). If all trials are included in the analysis, the results would be between an absolute risk increase of one per 1000 to 19 per 1000 participants.

There is an additional concern when interpreting our results of overdiagnosis. In cancer screening trials, persistent excess cancers in the screening arm is seen as evidence of overdiagnosis. For this to be true, the follow-up since the last screening round needs to be longer than the lead time, *i.e.* the average length

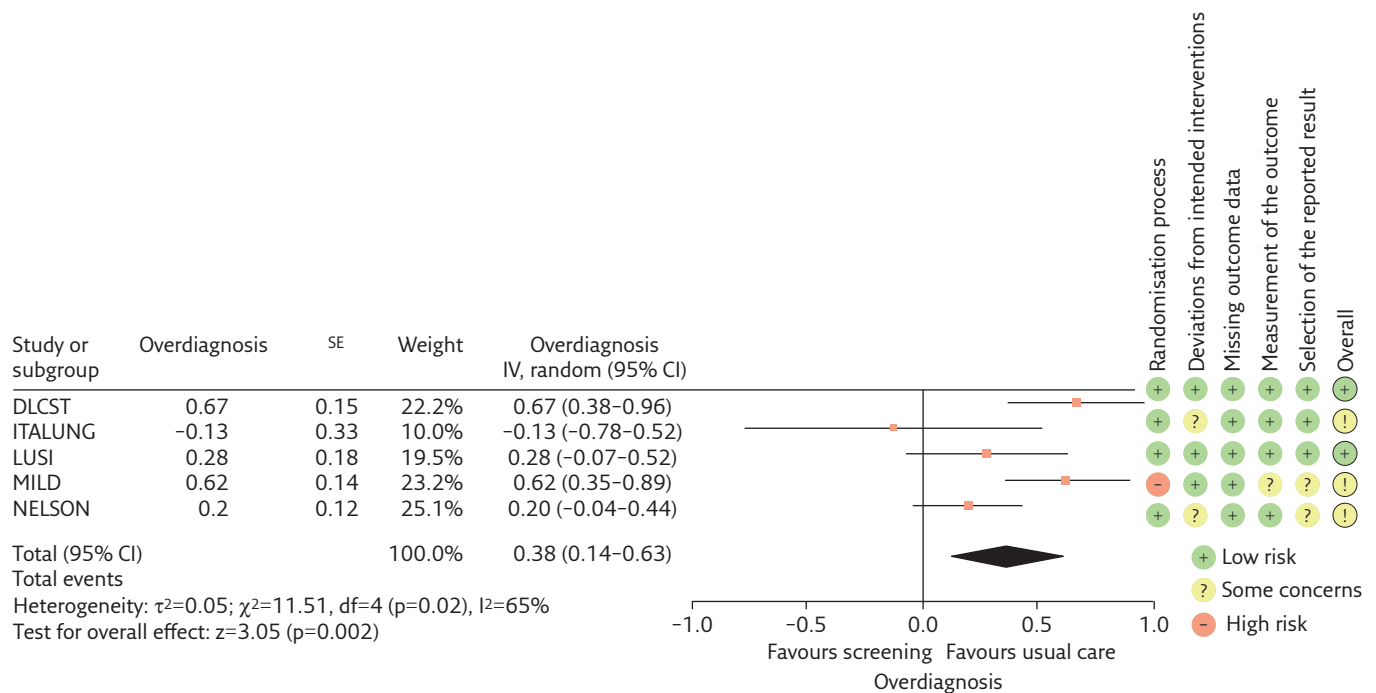


Figure 2 Forest plot of estimates of overdiagnosis defined as the fraction of screen-detected lung cancers that represent overdiagnosis. Meta-analysis includes all trials, regardless of bias assessment. Trials are listed alphabetically. IV: inverse variance; df: degrees of freedom.

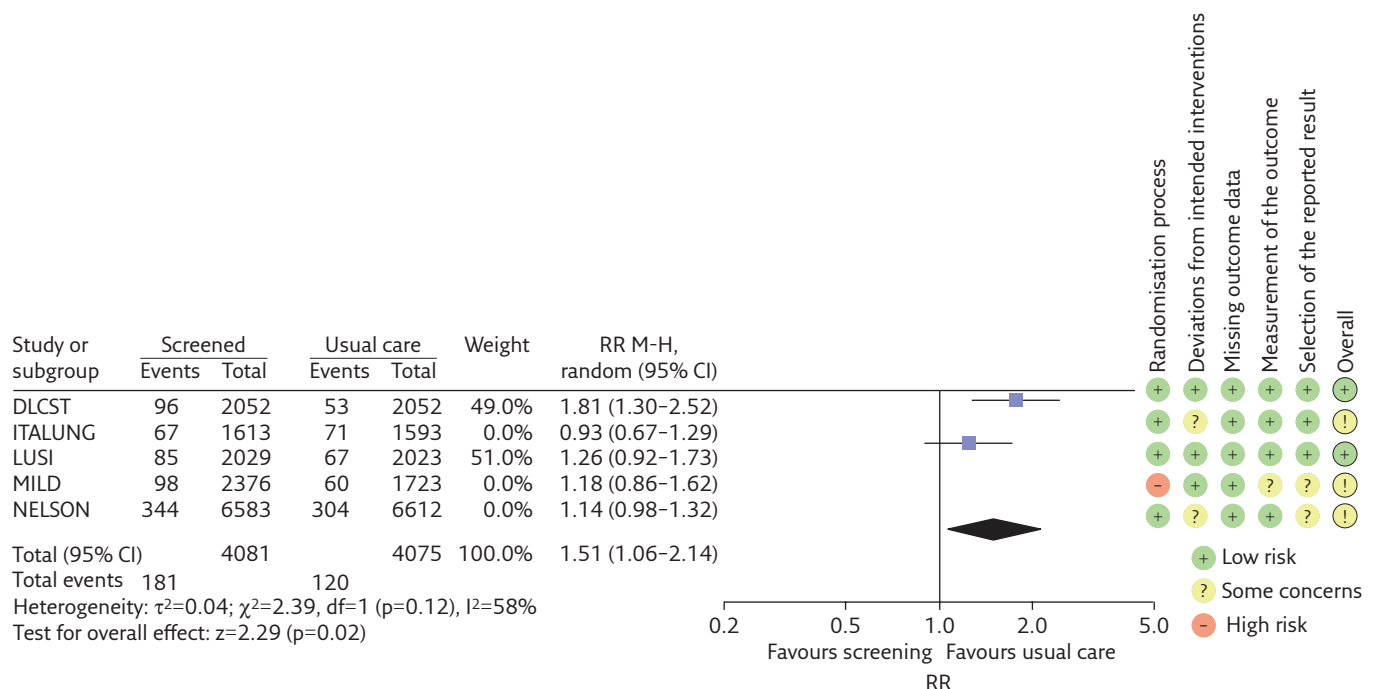


Figure 3 Forest plot of the RR of the cumulative incidence of lung cancer (estimates >1 represent overdiagnosis). The meta-analysis only includes low risk of bias trials. Trials are listed alphabetically. M-H: Mantel-Haenszel; df: degrees of freedom.

of time the diagnosis of a cancer is brought forward through screening [35]. The LUSI trial included in the meta-analysis has <3.6 years of follow-up since the last screening round [30]. In addition, the follow-up time in the MILD trial is unclear due to nonreporting of the screening-free interval for participants in both

the protocol and in publications [31, 36, 37]. Thus, it is possible that our current pooled estimate of the number of excess cancers in the screened group is affected by unaccounted lead time. If that is true, our results could overestimate the extent of overdiagnosis in lung cancer screening. Conversely, contamination

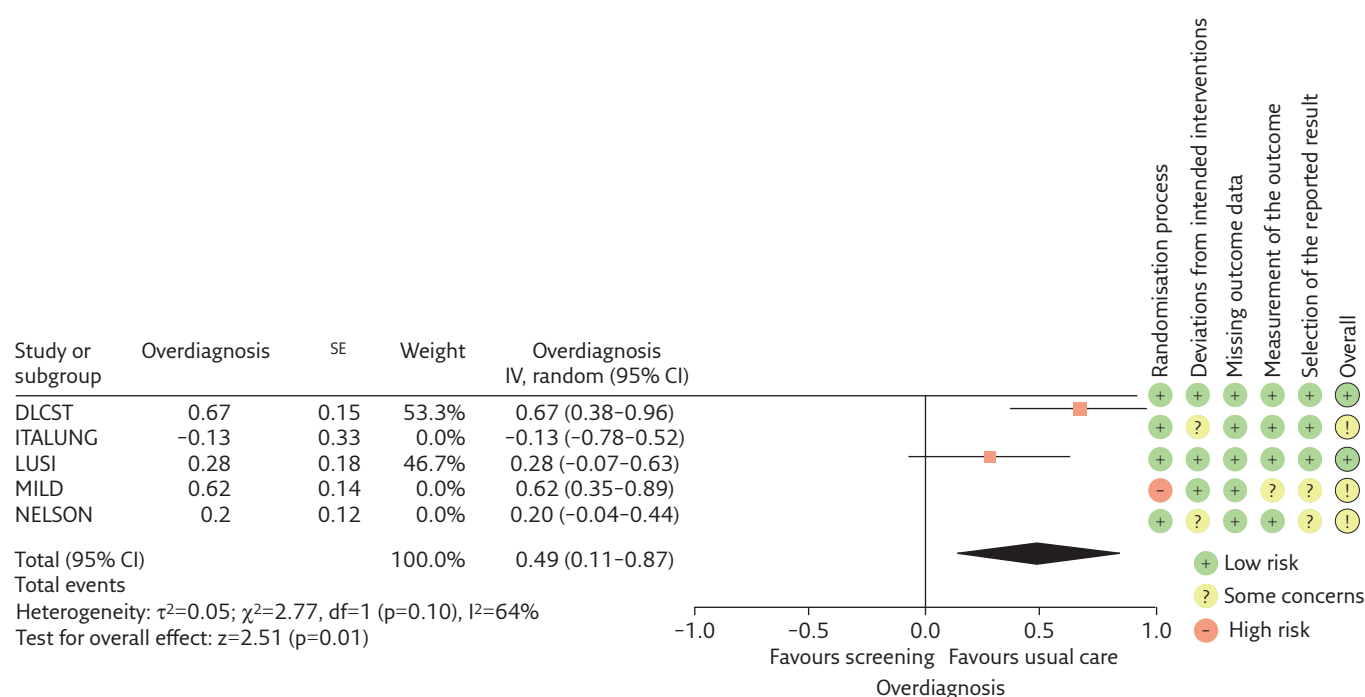


Figure 4 Forest plot of estimates of overdiagnosis defined as the fraction of screen-detected lung cancers that represent overdiagnosis. The meta-analysis only includes low risk of bias trials. Trials are listed alphabetically. IV: inverse variance; df: degrees of freedom.

of the control group would bias the estimate of overdiagnosis towards the null. Contamination of the control group was 11.4% in the LUSI trial and 1.2% in the MILD trial [22, 30]. Contamination was assessed using a questionnaire in the LUSI trial and, judging from the publication by phone calls, e-mail and contacts with general practitioners or referring hospitals in the MILD trial [22, 30]. The contamination rate is likely underreported in both LUSI and MILD because the methods used to estimate contamination rates in these two trials are not as valid the high-quality registries used in the DLCST, where contamination of the control group was 20.3% [29, 38].

Interpretation of the results in the context of previous literature

Other estimates of the effect of screening on long-term cancer incidence

The NLST, the previously mentioned, large US trial comparing screening for lung cancer with LDCT scans against screening with chest radiography, also found evidence of overdiagnosis [18]. They reported an 18.5% probability (95% CI 5.4%–30.6%) that any lung cancer detected by screening with LDCT was an overdiagnosis. In an 11.3-year follow-up of the NLST, this 18.5% probability of overdiagnosis had decreased to 0%. Whether this is due to an actual compensatory drop in the incident lung cancer rate or whether this is due to contamination of the control – or both – is uncertain [39]. The degree of contamination is not reported in the 11.3-year follow-up that went to the end of 2014

[39]. In 1996, the US Preventive Service Task Force (USPSTF) recommended against screening for lung cancer [40]. In 2004, the task force recommended neither for nor against [40], while in 2013, the USPSTF recommended annual screening with LDCT for lung cancer in people who had a smoking history of 30 pack-years and were aged 55–80 years [41]. The timing between the different USPSTF recommendations and the drop in the overdiagnosis estimate raises concern that contamination of the control group is most likely a part of the explanation. In support of this hypothesis, a theoretical lead-time of 3.6 years for lung cancer should lead to a compensatory drop in the incident lung cancer rate occurring sooner than after an 11.3-year follow-up.

Researchers from the NLST also used a convolution model to address two limitations of their study (the active comparator and the short follow-up period). They estimated that 11% of the screen-detected cancers could be due to overdiagnosis after accounting for model-based lead time [18]. Another modelling study estimated that overdiagnosis relative to no screening would be 10–12% of the screen-detected cancers across a wide range of simulated lung cancer screening programmes [42]. Of note, a systematic review found that no modelling studies on overdiagnosis in cancer screening relied on systematically reviewed evidence. In addition, the modelling studies did not validate their models using datasets from populations who were different from those in which the models were developed and calibrated. Finally, the included populations were different from those in which the models were calibrated. All this should cause doubt about the

validity of the findings from modelling studies on overdiagnosis in cancer screening [15].

Best estimates of the benefit of screening

A recent systematic review found that LDCT scan screening reduced lung cancer mortality (RR 0.83, 95% CI 0.76–0.90) [43]. This corresponds to an absolute risk reduction of seven lung cancer deaths per 1000 people screened, assuming a median baseline risk in the control group of 33 lung cancer deaths per 1000 participants. The systematic review did not find a significant effect on all-cause mortality (RR 0.95, 95% CI 0.90–1.00) [43]. Assuming the lung cancer mortality reduction is correct and assuming a median baseline risk in the control group of 114 overall deaths per 1000 participants, this corresponds to an absolute risk reduction of four deaths due to any cause per 1000 participants.

Clinical implications

Overdiagnosis is one of the most serious adverse effects of screening. In people with similar characteristics to those included in the European trials, we currently estimate an increase of 20 cancers per 1000 participants screened with LDCT scans. Based on recent systematic reviews, we expect that screening would prevent seven lung cancer deaths per 1000 participants, giving a ratio of approximately three overdiagnosed lung cancers per lung cancer death averted by LDCT screening. Besides this information, it should be communicated to eligible LDCT screening participants that there is a risk of a false-positive screening result. The average false-positive rate per screening round varies hugely in the RCTs: 23% in the NLST [17] and 3% in the DLCST [44], while in an ongoing LDCT screening programme, the false-positive rate might be as high as nearly 60% [45]. Moreover, there is evidence from all the European trials that LDCT screening for lung cancer does not reduce overall mortality.

There are two ongoing trials of screening for lung cancer with LDCT scans that have not published their final results: UKLS (4055 participants) [26] and the Chinese trial (6717 participants) [32]. It is possible that the estimate of overdiagnosis will change with the publication of these trials, and with the availability of more mature data from the LUSI and MILD trials [22, 30].

Implications for research

Lung cancer screening leads to a reduction in mortality while it also leads to additional harm due to overdiagnosis and a high number of false positives. Decision makers face a difficult situation trying to balance the benefits and harms of lung cancer screening: firstly, because the decision whether “benefit outweighs harms” is value laden, and likely varies between individuals and

Self-evaluation questions

1. What are the three main reasons that lead to overdiagnosis in cancer imaging screening?
2. Besides the usual biases that are supposed to be assessed in randomised controlled trials, which two additional biases are of importance when estimating the degree of overdiagnosis in randomised controlled cancer screening trials?
3. What is the extent of overdiagnosis in lung cancer screening with low-dose computed tomography scans?

across cultural settings; and secondly, because tools to assist decision makers with balancing the benefits and harms of screening are sparse [45, 46]. Moreover, the balance between benefits and harms of LDCT lung cancer screening should be viewed with its direct and indirect costs to individuals and society compared to public health alternatives such as primary tobacco prevention. Researchers need to assess potential ways to minimise harms of LDCT screening, especially the degree of overdiagnosis. The source of the variation in lung cancer overdiagnosis rates is currently unknown but should be the focus of future research to better understand what causes it and how to prevent it if lung screening programmes are implemented. Sources of possible variation that could be investigated include:

- variation in the population screened (age group screened, definition of heavy smokers and other risks specific to the screened population including, *e.g.* asbestos and genomic variability)
- differences in screening practice (screening intervals, numbers of screening rounds, differences in screening technology and differences in how abnormal findings are managed)
- professional differences (radiological and pathological thresholds)

A first potential strategy to reduce overdiagnosis is changing the eligibility criteria for LDCT screening. To date, participants have been selected based on age and smoking history. Now, there are risk prediction models that consider other variables, and may be better at increasing the benefits and decreasing the harms of LDCT screening. However, high-quality RCTs are needed to give valid answers to such research questions. The second strategy is changing the frequency of screening, the criteria for recall to further investigation, and thresholds to start or to stop screening. Modelling studies are exploring the consequences of changing screening parameters. Again, high quality evidence from RCTs is needed. Given the trade-offs between benefits and harms, researchers need to consider how decision aids may be incorporated into screening programmes to help healthy heavy smokers (current and former) make an informed choice about whether to participate in CT screening for lung cancer.

Conclusion

We found that screening increases the long-term, cumulative incidence of lung cancer (1.51, 95% CI 1.06–2.14) and that 49% of screen-detected cancers may be overdiagnosed (95% CI 11%–87%). This corresponds to an absolute risk increase of 20 cancers per 1000 people screened

(range 2–46 per 1000 screened), assuming a median baseline risk in the control group of 40 lung cancers per 1000 participants. However, there is great uncertainty about the degree of overdiagnosis in lung cancer screening due to the heterogeneity of all included trials, imprecision of the point estimate and large variation in the quality of the included trials.

Affiliations

John Brodersen^{1,2}, Theis Voss^{1,2}, Frederik Martiny^{1,2}, Volkert Siersma¹, Alexandra Barratt³, Bruno Heleno⁴

¹The Section of General Practice and the Research Unit for General Practice, Dept of Public Health, University of Copenhagen, Copenhagen, Denmark. ²The Research Unit for General Practice in Region Zealand, Sorø, Denmark. ³Sydney School of Public Health, University of Sydney, Sydney, Australia. ⁴CEDOC, Chronic Diseases Research Centre, NOVA Medical School, Faculdade de Ciências Médicas, Universidade NOVA de Lisboa, Lisbon, Portugal.

Supplementary material

This article has supplementary material available from breath.ersjournals.com.

Author contributions

J. Brodersen, T. Voss and B. Heleno conceived and designed this meta-analysis. J. Brodersen, T. Voss, F. Martiny and B. Heleno acquired, analysed and interpreted the data. J. Brodersen and B. Heleno drafted the manuscript. J. Brodersen, T. Voss, F. Martiny, V. Siersma, A. Barratt and B. Heleno critically revised the manuscript for important intellectual content.

Conflict of interest:

J. Brodersen is a member of the Danish Lung Cancer Screening Trial steering committee. Besides that, he has nothing to disclose. T. Voss has nothing to disclose. F. Martiny has nothing to disclose. V. Siersma has nothing to disclose. A. Barratt has nothing to disclose. B. Heleno has nothing to disclose.

References

- Brodersen J, Schwartz LM, Heneghan C, *et al*. Overdiagnosis: what it is and what it isn't. *BMJ Evid Based Med* 2018; 23: 1–3.
- Brodersen J, Schwartz LM, Woloshin S. Overdiagnosis: how cancer screening can turn indolent pathology into illness. *APMIS* 2014; 122: 683–689.
- Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010; 102: 605–613.
- Lindell RM, Hartman TE, Swensen SJ, *et al*. 5-year lung cancer screening experience: growth curves of 18 lung cancers compared to histologic type, CT attenuation, stage, survival, and size. *Chest* 2009; 136: 1586–1595.
- Ko JP, Berman EJ, Kaur M, *et al*. Pulmonary nodules: growth rate assessment in patients by using serial CT and three-dimensional volumetry. *Radiology* 2012; 262: 662–671.
- Wang J, Mahasittiwat P, Wong KK, *et al*. Natural growth and disease progression of non-small cell lung cancer evaluated with 18F-fluorodeoxyglucose PET/CT. *Lung Cancer* 2012; 78: 51–56.
- Fontana RS, Sanderson DR, Woolner LB, *et al*. Screening for lung cancer. A critique of the Mayo Lung Project. *Cancer* 1991; 67: Suppl. 4, 1155–1164.
- Harris RP, Sheridan SL, Lewis CL, *et al*. The harms of screening: a proposed taxonomy and application to lung cancer screening. *JAMA Intern Med* 2014; 174: 281–285.
- Welch HG, Schwartz L, Woloshin S. *Overdiagnosed: Making People Sick in the Pursuit of Health*. Boston, Beacon Press, 2011.
- Heleno B, Thomsen MF, Rodrigues DS, *et al*. Quantification of harms in cancer screening trials: literature review. *BMJ* 2013; 347: f5334.
- Harris RP, Wilt TJ, Qaseem A. A value framework for cancer screening: advice for high-value care from the American College of Physicians. *Ann Intern Med* 2015; 162: 712–717.
- Brodersen J, Jorgensen KJ, Gotzsche PC. The benefits and harms of screening for cancer with a focus on breast screening. *Pol Arch Med Wewn* 2010; 120: 89–94.
- Ostero J, Siersma V, Brodersen J. Breast cancer screening implementation and reassurance. *Eur J Public Health* 2014; 24: 258–263.
- Rasmussen JF, Siersma V, Pedersen JH, *et al*. Psychosocial consequences in the Danish randomised controlled lung cancer screening trial (DLCST). *Lung Cancer* 2015; 87: 65–72.
- Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ* 2015; 350: g7773.
- Patz EF, Jr., Pinsky P, Gatsonis C, *et al*. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2014; 174: 269–274.
- Aberle DR, Adams AM, Berg CD, *et al*. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365: 395–409.

18. Patz EF, Jr, Pinsky P, Gatsonis C. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2013; 174: 269–274.
19. Field JK, van Klaveren R, Pedersen JH, *et al.* European randomized lung cancer screening trials: post NLST. *J Surg Oncol* 2013; 108: 280–286.
20. Pedersen JH, Ashraf H, Dirksen A, *et al.* The Danish Randomized Lung Cancer CT Screening Trial – overall design and results of the prevalence round. *J Thorac Oncol* 2009; 4: 608–614.
21. Becker N, Motsch E, Gross ML, *et al.* Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. *J Cancer Res Clin Oncol* 2012; 138: 1475–1486.
23. Paci E, Puliti D, Lopes Pegna A, *et al.* Mortality, survival and incidence rates in the ITALUNG randomised lung cancer screening trial. *Thorax* 2017; 72: 825–831.
24. Infante M, Lutman FR, Cavuto S, *et al.* Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer* 2008; 59: 355–363.
25. van Klaveren RJ, Oudkerk M, Prokop M, *et al.* Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361: 2221–2229.
26. Baldwin DR, Duffy SW, Wald NJ, *et al.* UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. *Thorax* 2011; 66: 308–313.
27. Sterne JAC, Savovic J, Page MJ, *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019; 366: l4898.
28. Marmot MG, Altman DG, Cameron DA, *et al.* The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013; 108: 2205–2240.
29. Heleno B, Siersma V, Brodersen J. Estimation of overdiagnosis of lung cancer in low-dose computed tomography screening. *JAMA Intern Med* 2018; 178: 1420–1422.
30. Becker N, Motsch E, Trotter A, *et al.* Lung cancer mortality reduction by LDCT screening—Results from the randomized German LUSI trial. *Int J Cancer* 2019; 146: 1503–1513.
31. Pastorino U, Silva M, Sestini S, *et al.* Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. *Ann Oncol* 2019; 30: 1162–1169.
32. Yang W, Qian F, Teng J, *et al.* Community-based lung cancer screening with low-dose CT in China: Results of the baseline screening. *Lung Cancer* 2018; 117: 20–26.
33. de Koning HJ, van der Aalst CM, de Jong PA, *et al.* Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med* 2020; 382: 503–513.
34. Baecke E, de Koning HJ, Otto SJ, *et al.* Limited contamination in the Dutch-Belgian randomized lung cancer screening trial (NELSON). *Lung Cancer* 2010; 69: 66–70.
35. Zahl PH, Jorgensen KJ, Gotzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. *Br J Cancer* 2013; 109: 2014–2019.
36. Pastorino U, Rossi M, Rosato V, *et al.* Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev* 2012; 21: 308–315.
37. Pastorino U. Early Lung Cancer Detection in High Risk Individuals (MILD). <https://clinicaltrials.gov/ct2/show/record/NCT02837809>.
38. Saghir Z, Ashraf H, Dirksen A, *et al.* Contamination during 4 years of annual CT screening in the Danish Lung Cancer Screening Trial (DLCST). *Lung Cancer* 2011; 71: 323–327.
39. The National Lung Screening Trial Research Team. Lung Cancer Incidence and Mortality with Extended Follow-up in the National Lung Screening Trial. *J Thorac Oncol* 2019; 14: 1732–1742.
40. USPSTF. Lung cancer screening: recommendation statement. *Ann Intern Med* 2004; 140: 738–739.
41. Moyer VA. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014; 160: 330–338.
42. de Koning HJ, Meza R, Plevritis SK, *et al.* Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med* 2014; 160: 311–320.
43. Huang K-L, Wang S-Y, Lu W-C, *et al.* Effects of low-dose computed tomography on lung cancer screening: a systematic review, meta-analysis, and trial sequential analysis. *BMC Pulm Med* 2019; 19: 126.
44. Saghir Z, Dirksen A, Ashraf H, *et al.* CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; 67: 296–301.
45. Kinsinger LS, Anderson C, Kim J, *et al.* Implementation of lung cancer screening in the Veterans Health Administration. *JAMA Intern Med* 2017; 177: 399–406.
46. Caverly TJ, Fagerlin A, Wiener RS, *et al.* Comparison of observed harms and expected mortality benefit for persons in the Veterans Health Affairs Lung Cancer Screening Demonstration Project. *JAMA Intern Med* 2018; 178: 426–428.